SODALI &CO | HARBOTTLE & LEWIS | LevelBlue

THE AI-ENABLED THREAT LANDSCAPE:

REAL WORLD LESSONS FROM LAWYERS, PR, AND CYBER SECURITY EXPERTS

AI is changing the landscape for managing cyber security risks – affecting attackers, defenders, and responders alike.

The last decade has witnessed the transition of cyber security threats from a technical challenge to one of the key drivers of business and operational risk globally. Now comes the rapid evolution of artificial intelligence (AI) to change the game once again. In 2026, AI in the context of cyber security is multifaceted: a tool for defence and incident response – but also a powerful accelerator for attackers. We are all witnessing a rapidly changing landscape where AI has started to enable faster, cheaper, more scalable, and highly personalised cyber incidents. It has also significantly lowered the bar of entry for threat actors and is a potent tool for financially motivated crime and hacktivism, not just sophisticated nation-states.

Once again, Boards must rapidly get up to speed with this new AI-enabled risk landscape. They can start with better understanding these fast evolving AI-driven threats, as they can lead to material financial losses from operational disruptions, regulatory penalties and longer tail reputational damages.

This educational overview is designed for a general, non-technical audience of business leaders and Board members. Its aim is to help build awareness of the main categories of AI-enabled cyber threats and provide key talking points for discussions with risk, legal, communications, and cyber security teams supporting the Board and the wider organisation.

THE AI ATTACK LANDSCAPE: 2026 TAXONOMY

The following table explores a wide range of new or significantly evolved AI-enabled cyber-attacks that have emerged or gained serious traction based on current threat intelligence reporting.

AI RISK	EXPLANATION
AI-ORCHESTRATED AUTONOMOUS CYBER-ATTACKS	Nation state sponsored actors have been observed to manipulate agentic large language models (LLMs) to handle the entire attack: scouting targets, finding weak spots, breaking in, stealing passwords, moving deeper into networks, grabbing sensitive data and sending it back to the threat actor. Although fully autonomous AI-driven attacks have not yet been observed in the wild, emerging agentic LLM capabilities are already enabling threat actors to automate stages of cyber-attacks and to expand the scale of the attacks.[1,2] Current research into a models' ability to conduct aspects of the attacks show this is an emerging threat.[3]
EXPANDED ATTACK SURFACE & EXPLOITATION OF TRUST	Attackers are targeting enterprise AI tools and workflows, expanding the attack surface. Recent reporting found that threat actors are exploiting legitimate AI tools by injecting malicious prompts, exploiting vulnerabilities in AI development platforms, and publishing malicious AI servers impersonating trusted services to intercept sensitive data.[4] Using advanced techniques such as prompt injection, threat actors can hide malicious commands inside normal-looking files – an email attachment, calendar invite, photo, or document – that an AI tool reads as part of its job. The hidden text tricks the AI into ignoring its rules: leaking secrets, writing dangerous code, or misusing connected tools.
AI-ENHANCED PHISHING & SOCIAL ENGINEERING	AI can study a colleague's emails, posts, or writing style, then write fake messages that sound exactly like their work colleagues. Moreover, AI can reply in real time, which can continue the momentum of the scam.[5] AI has been used extensively by nation-state threat actors to enhance credibility of false applicants for IT-related jobs.[6] Ransomware operators have turned to AI chatbots to make ransom demand negotiations more efficient and apply greater pressure on victims.[7]
"JAILBREAKING" COMMERCIAL AI TOOLS	There is ongoing work by AI tool developers to build in safety and security features to prevent misuse in cyber-attacks, which are being challenged in real time by cyber threat actors in various ways. Threat actors have been identifying means to "jailbreak" the models to provide output to further various stages of cyber-attacks – such as by convincing the AI to assist with purportedly legitimate activity such as penetration testing, bug bounty programs, and "capture the flag" challenges.[8,9]
POISONING ATTACKS ON AI TRAINING DATA	Threat actors can attempt to "poison" LLMs simply by injecting specific text into public content to make a model learn undesirable or dangerous behaviour. Using poisoning techniques such as "backdoors", the AI behaves normally until it sees a secret trigger phrase – then it flips: starts leaking data, ignores safety rules, or helps attackers. There is evidence that this can be done successfully with even a relatively small amount of malicious documents.[10]
AI-ASSISTED MALWARE DEVELOPMENT	Malicious actors may use tools to generate or rapidly modify malware code, allowing malicious software to be quickly adapted or modified to better evade security controls[11]
DEEPFAKE-ENABLED MISINFORMATION & FRAUD	AI can create ultra-realistic fake video calls or voice messages of your CEO, CFO, or family member in real time. Criminals use them for high-stakes scams e.g. wire fraud scenarios.[12,13] Deepfakes and other synthetic, AI-enabled media can pose challenges for authentication and attribution.[14]
LOWER BARRIERS & UPSKILL NEW THREAT ACTORS	AI tools are reducing the technical expertise required to conduct cyber operations, potentially enabling a broader range of actors, including less sophisticated group types such as hacktivists or financially motivated threat actors.[15] The cyber-crime marketplace has expanded to offer AI tooling to unsophisticated threat actors to support phishing, ransomware, and vulnerability identification.[16]

KEY TALKING POINTS FOR BOARD PRESENTATIONS

We've reframed the talking points to focus on the legal, reputational, and cyber security consequences of AI risks for the organisation, making them more accessible and actionable for the Board.

This approach redirects conversations from "how the technology operates" to "what it means for the company," enabling focused discussions on: governance, liability, and organisational resilience.

1. Enterprise Risk

- Attackers now move dramatically faster – traditional perimeter defences can be overwhelmed by AI-accelerated speed and scale. This compresses detection and response windows to near-zero, raising the likelihood of widespread compromise before intervention.
- Organisation's own AI tools expand the attack surface – internal generative AI systems and models become new vectors, vulnerable to prompt injection, data poisoning, model theft, or adversarial inputs that can leak sensitive information or produce harmful outputs.
- AI-assisted defensive tools are improving the ability to detect and block attacks by analysing large volumes of security data to quickly identify suspicious behavioural patterns, enabling faster and more automated reposes to threats.

2. Legal & Regulatory Exposure

- The UK's Cyber Security and Resilience Bill (introduced Nov 2025), which modernises the UK's implementation of the Network and Information Systems (NIS) 2018 regulation, will – as parts of its reforms – have far-reaching implications for AI deployment, requiring safe, reliable AI usage and robust, evidence-based security measures along with enhanced reporting and notification requirements.
- Whilst companies should implement AI usage policies and educate employees, directors must demonstrate active oversight of AI risks as part of enterprise risk management. Failure to do so risks shareholder derivative suits, scrutiny during activism, or personal Directors & Officers (D&O) liabilities.

3. Reputation Management

- Deepfakes and hyper-personalised scams destroy trust at scale – a convincing fake CEO video can go viral in hours or voice authorising fraudulent wire transfers can lead to immediate financial loss, and both can trigger media firestorms, and eroded stakeholder confidence.
- High profile AI failures – such as biased or unfair decision making, hallucinations, or privacy breaches – can rapidly gain attention across social platforms, leading to customer loss, partner hesitation, and long term brand damage. Market reactions are often swift and severe, unless strong preparedness and response measures are firmly in place.

ACTION POINTS

1. Enterprise Risk

UPDATE INCIDENT RESPONSE PLANS & PLAYBOOKS	AI-enabled cyber-attacks introduce a host of changes, which should be reflected in the advance planning processes. Consider AI-attack tabletop exercises at the Board level and build the tactics into business-as-usual (BAU) testing, such as red teaming. Ensure that your incident responders are equipped to preserve and investigate the unique types of evidence involved in such attacks, and structure those engagements to help protect claims of legal privilege.
OVERSEE MANDATORY AI-SOCIAL ENGINEERING TRAINING	Ensure there is regular, realistic training and simulations on deepfakes, voice cloning, and video or audio impersonation. Review completion rates, phishing test metrics, and social-engineering incident trends.
APPROVE & MONITOR AI-ENHANCED SECURITY INVESTMENTS	Consider adopting AI-driven detection, response, and prevention tools that match or exceed adversary speed and sophistication.
REQUIRE SECURE-BY-DESIGN FOR ALL AI INITIATIVES	Mandate vulnerability testing, access controls, and lifecycle risk assessments for internal development, third-party generative tools, and embedded AI models.
ESTABLISH BOARD-OVERSEEN AI GOVERNANCE FRAMEWORK	Approve formal AI policies, define risk tolerance thresholds, assign monitoring responsibilities, and designate oversight. Demand regular, structured Board reporting on AI usage, incidents, and compliance.
ENSURE DYNAMIC THREAT INTELLIGENCE INTEGRATION	Use relevant intelligence on emerging AI-enabled threats to ensure security controls reflect emerging AI-enabled threats.

2. Legal & Regulatory Exposure

APPOINT THE RIGHT LEGAL EXPERTS TO TURN TO AT VERY SHORT NOTICE AHEAD OF A CRISIS	Fake viral content or online, emerging attacks must be stopped at speed. Minutes matter, so companies must make sure that they know which legal experts to urgently turn to. Appoint experts in advance and pre-prepare template legal correspondence to deploy urgently in a crisis.
LAW IS PART OF, NOT THE WHOLE SOLUTION	Make sure that your legal team is joined up with your other advisors in this area, such as digitial forensics and communications. Collaboration is key when dealing with an AI driven threat, such as a deep fake or fast moving mis-information campaign, particularly as press briefings to change the narrative of adverse fake content, based upon a credible forensic determination that content is fake, will be just as essential to persuading the media not to publish or broadcast it in order to protect client reputation.

CHALLENGE FAKE CONTENT, BOTH PRE & POST PUBLICATION OR BROADCAST	As soon as stolen or fake online or false social media content is detected via monitoring services, consider using legal tools or platform terms and conditions to get it removed. Injunctions may also be available. Otherwise, such content will be picked up in the mainstream and other media (particularly if it is being amplified by malicious sources). If journalists get hold of the fake story and come for comment pre-publication, engage legal tools to explain and try and stop the story. If the story is published without warning, use the same legal tools to have the story amended or removed accordingly. Consider removing URLs to stories from online search engines, AI chatbot tools, LLMs and due diligence platforms, otherwise the fake content will live online forever and could resurface as part of other stories/affect future business relationships.
AI POLICIES & EMPLOYEES	Enact appropriate and enforceable AI policies for employees, take employment law advice on obligations and on enforcing breaches of the policy, educate employees. Use copyright law in relation to AI generated deepfakes created by employees in breach to remove content and stop republications or dissemination.
PROTECTION VIA CONTRACTUAL OBLIGATIONS OF MANAGED SERVICE PROVIDERS (MSPS)	If the responsibility for maintaining the security of your IT estate falls upon MSPs, carry out a contractual review of their obligations in relation to cyber attacks and AI generally and, if possible, re-negotiate to strengthen your position.
DATA AUDITS	Carry out a data audit in order to understand the data map of all the processing taking place within your organisation, what you hold and where, along with standard UK General Data Protection Regulation (GDPR) compliance reviews and assessments.
CARRY OUT A JOINT, LEGAL & TECHNICAL, BREACH SIMULATION EXERCISE	Test the resilience of your crisis and cyber response plans by carrying out a breach simulation exercise with legal and forensic experts and strengthen defences following the exercise. Continue to educate all employees and and ask your legal team to update you on new laws and regulations and/or attend their seminars and webinars.

3. Reputation Management

DEVELOP A DEDICATED AI-CRISIS PLAYBOOK	Ensure your organisation has a specific framework for AI threats, including verification protocols, statement frameworks, and rapid response templates to ensure quick, consistent action during deepfake or misinformation incidents.
ENSURE MONITORING SYSTEMS ARE IN PLACE TO SPOT TROUBLE EARLY	Consider using AI-driven tools and social listening platforms to watch for weak spots, monitor for emerging threats, and detect anomalies like sudden spikes in disinformation in real-time.
BOARD LEVEL ASSURANCE ON AI CRISIS PREPAREDNESS DRILLS	Champion the importance of tabletop exercises for communications and executive teams, covering AI literacy, deepfake recognition, and rapid-response tactics. Each exercise should stress test the crisis playbook by exploring realistic worst-case scenarios to strengthen cross functional coordination.
SUPPORT AI LITERACY ACROSS THE BOARD TO STRENGTHEN GOVERNANCE	In the baseline scenario, this is typically achieved through targeted training for directors and regular management reporting on AI use. For companies with significant AI exposure - such as AI developers or organisations embedding AI into core operations - Boards should also consider recruiting candidates with deeper, specific AI expertise. This helps ensure the Board can better understand and oversee the operational, ethical, and strategic implications of AI, leading to more informed decision-making and effective risk management.

SOURCES

1. Disrupting the first reported AI-orchestrated cyber espionage campaign \ Anthropic
2. AI-Powered Hacker Steals 150GB from Mexican Government Using Anthropic's Claude
3. Cyber Model Arena | Wiz
4. CrowdStrike 2026 Global Threat Report: Evasive Adversary Wields AI
5. How to Tackle Evolving Email-Based Attacks
6. North Korean agents using AI to trick western firms into hiring them, Microsoft says | Technology sector | The Guardian
7. Ransomware Group Uses AI Chatbot to Intensify Pressure on Victims – Infosecurity Magazine
8. Detecting and countering misuse of AI: August 2025 \ Anthropic
9. GTIG AI Threat Tracker: Advances in Threat Actor Usage of AI Tools | Google Cloud Blog
10. A small number of samples can poison LLMs of any size \ Anthropic
11. GTIG AI Threat Tracker: Advances in Threat Actor Usage of AI Tools | Google Cloud Blog
12. Arup Deepfake Scam: How $25M Was Stolen via Video Call
13. What CIOs Can Learn from an Attempted Deepfake Call
14. Deepfake Media Forensics: Status and Future Challenges – PMC
15. AI as tradecraft: How threat actors operationalize AI | Microsoft Security Blog
16. GTIG AI Threat Tracker: Advances in Threat Actor Usage of AI Tools | Google Cloud Blog

KEY CONTACTS

For more information please contact:

Michael Yates
Partner, Harbottle & Lewis
michael.yates@harbottle.com

Brandy Wityak
VP, Level Blue
brandy.wityak@levelblue.com

Robert Probin
Manager, Level Blue
robert.probin@levelblue.com

Jenny Pirault
Manager, Sodali & Co
jenny.pirault@sodali.com